

On Regularizing Effects of MINRES and MR-II for Large Scale Symmetric Discrete Ill-posed Problems*

Yi Huang^a, Zhongxiao Jia^b

^a Department of Mathematical Sciences, Tsinghua University, 100084 Beijing, China

^b Department of Mathematical Sciences, Tsinghua University, 100084 Beijing, China,

E-mail: huangyi10@mails.tsinghua.edu.cn, jiazx@tsinghua.edu.cn

Abstract

For large scale symmetric discrete ill-posed problems, MINRES and MR-II are often used iterative regularization solvers. We call a regularized solution best possible if it is at least as accurate as the best regularized solution obtained by the truncated singular value decomposition (TSVD) method. In this paper, we analyze their regularizing effects and establish the following results: (i) the filtered SVD expression are derived for the regularized solutions by MINRES; (ii) a hybrid MINRES that uses explicit regularization within projected problems is needed to compute a best possible regularized solution to a given ill-posed problem; (iii) the k th iterate by MINRES is more accurate than the $(k - 1)$ th iterate by MR-II until the semi-convergence of MINRES, but MR-II has globally better regularizing effects than MINRES; (iv) bounds are obtained for the 2-norm distance between an underlying k -dimensional Krylov subspace and the k -dimensional dominant eigenspace. They show that MR-II has better regularizing effects for severely and moderately ill-posed problems than for mildly ill-posed problems, and a hybrid MR-II is needed to get a best possible regularized solution for mildly ill-posed problems; (v) bounds are derived for the entries generated by the symmetric Lanczos process that MR-II is based on, showing how fast they decay. Numerical experiments confirm our assertions. Stronger than our theory, the regularizing effects of MR-II are experimentally shown to be good enough to obtain best possible regularized solutions for severely and moderately ill-posed problems.

Keywords: Symmetric ill-posed problem, regularization, partial regularization, full regularization, semi-convergence, MR-II, MINRES, LSQR, hybrid.

Mathematics Subject Classifications (2010): 65F22, 65J20, 15A18.

1 Introduction

Consider the large scale discrete linear ill-posed problem

$$Ax = b, \quad A \in \mathbb{R}^{n \times n}, \quad b \in \mathbb{R}^n, \quad (1.1)$$

where A is symmetric and extremely ill conditioned with its singular values decaying gradually to zero without a noticeable gap. This kind of problem arises in many science and engineering areas [10]. The right-hand side b is noisy and typically affected by a white noise, caused by measurement, modeling or discretization errors, i.e.,

$$b = \hat{b} + e,$$

*This work was supported in part by the National Science Foundation of China (No. 11371219).

where $e \in \mathbb{R}^n$ represents a white noise vector and $\hat{b} \in \mathbb{R}^n$ denotes the noise-free right-hand side, and it is supposed that $\|e\| < \|\hat{b}\|$. Because of the presence of noise e in b and the high ill-conditioning of A , the naive solution $x_{naive} = A^{-1}b$ of (1.1) is far from the true solution $x_{true} = A^{-1}\hat{b}$ and meaningless. Therefore, one needs to use regularization to determine a regularized solution so that it is close to $x_{true} = A^{-1}\hat{b}$ as much as possible [8, 10].

For A symmetric, its SVD is closely related to its spectral decomposition as follows:

$$A = V\Lambda V^T = V\Omega\Sigma V^T = U\Sigma V^T, \quad (1.2)$$

where $U = (u_1, u_2, \dots, u_n) = V\Omega$ and $V = (v_1, v_2, \dots, v_n)$ are orthogonal, whose columns are the left and right singular vectors of A , respectively, the diagonal matrix $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ with the singular values labeled as $\sigma_1 > \sigma_2 > \dots > \sigma_n > 0$, $\Omega = \text{diag}(\pm 1)$ is a signature matrix such that $\sigma_i = |\lambda_i|$ with the λ_i the eigenvalues of A , and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Obviously, $u_i = \pm v_i$ with the \pm sign depending on Ω . With (1.2), we can express the naive solution of (1.1) as

$$x_{naive} = \sum_{i=1}^n \frac{v_i^T b}{\lambda_i} v_i = \sum_{i=1}^n \frac{v_i^T \hat{b}}{\lambda_i} v_i + \sum_{i=1}^n \frac{v_i^T e}{\lambda_i} v_i = x_{true} + \sum_{i=1}^n \frac{v_i^T e}{\lambda_i} v_i. \quad (1.3)$$

Throughout the paper, we assume that \hat{b} satisfies the discrete Picard condition [8, 10]: On average, the coefficients $|u_i^T \hat{b}| = |v_i^T \hat{b}|$ decay faster than the singular values $\sigma_i = |\lambda_i|$. This is a necessary hypothesis that controls the size of $\|x_{true}\|$ and makes regularization possible to find meaningful approximations to x_{true} [8, 10]. For the sake of simplicity, we assume that they satisfy a widely used model [8, 10]:

$$|v_i^T \hat{b}| = |\lambda_i|^{1+\beta}, \quad \beta > 0, \quad i = 1, 2, \dots, n. \quad (1.4)$$

Similar to the truncated SVD (TSVD) method [8, 10], for A symmetric, a truncated spectral decomposition method obtains the TSVD regularized solutions

$$x_k = \sum_{i=1}^k \frac{v_i^T b}{\lambda_i} v_i = \sum_{i=1}^k \frac{u_i^T b}{\sigma_i} v_i = A_k^\dagger b, \quad k = 1, 2, \dots, n, \quad (1.5)$$

where $A_k = U_k \Sigma_k V_k^T$ with U_k and V_k the first k columns of U and V , respectively, $\Sigma_k = \text{diag}(\sigma_1, \dots, \sigma_k)$, and † the Moore-Penrose generalized inverse of a matrix. Obviously, x_k is the minimum 2-norm least squares solution of the perturbed problem that replaces A in (1.1) by its best rank k approximation A_k .

Let k_0 denote the transition point such that $|v_{k_0}^T \hat{b}| > |v_{k_0+1}^T e|$ and $|v_{k_0+1}^T \hat{b}| \approx |v_{k_0+1}^T e|$, which divides the eigenvalues λ_i or equivalently the singular values $\sigma_i = |\lambda_i|$ into the dominant or large ones for $i \leq k_0$ and the small ones for $i > k_0$. It is known from [8, p. 176] and [10, p. 86-88] that the best TSVD regularized solution is x_{k_0} , which consists of the k_0 dominant SVD components of A , i.e., the dominant spectral components corresponding to the first k_0 large eigenvalues in magnitude when A is symmetric. A number of approaches have been proposed for determining k_0 , such as discrepancy principle, the discrete L-curve criterion and the generalized cross validation (GCV); see, e.g., [1, 4, 8, 15, 21] for comparisons of the classical and new ones. In our numerical experiments, we do this using the L-curve criterion in the TSVD method and Krylov iterative methods.

The TSVD method is important in its own right and plays a central role in analyzing the standard-form Tikhonov regularization [8, 10]. It and the standard-form Tikhonov regularization expand their solutions in the basis vectors v_i and produce very similar solutions with essentially the minimum 2-norm error; see [8, p. 109-111] and [10, Sections 4.2 and 4.4]. Therefore, the TSVD

method can get a best regularized solution to (1.1), and it has long been used as a general-purpose reliable and efficient numerical method for solving a small and/or moderate sized (1.1) [8, 10]. As a result, we will take the TSVD solution x_{k_0} as a standard reference when assessing the regularizing effects of iterative solvers and accuracy of iterates under consideration in this paper.

For (1.1) is large, it is generally impractical to compute the spectral decomposition of A . In this case, one typically solves it iteratively via some Krylov subspace methods [8, 10]. For (1.1) with a general matrix A , the mathematically equivalent LSQR [19] and CGLS [2] have been most commonly used for years and have been shown to have intrinsic regularizing effects [6, 8, 10]. They exhibit the semi-convergence; see [6], [8, p. 135], [10, p. 110]: The iterates tend to be better and better approximations to the exact solution x_{true} and their norms increase slowly and the residual norms decrease. In later stages, however, the noise e starts to deteriorate the iterates, so that they will start to diverge from x_{true} and instead converge to x_{naive} , while their norms increase considerably and the residual norms stabilize. For LSQR, the semi-convergence is due to the fact that the projected problem at some iteration starts to inherit the ill-conditioning of (1.1), that is, the noise progressively enters the solution subspace, so that a small singular value of the projected problem appears and the regularized solution is deteriorated [8, 10].

As far as an iterative solver for solving (1.1) is concerned, a central problem is whether or not it has already obtained a best possible regularized solution at semi-convergence. Here, as defined in the abstract, a best possible regularized solution means that it is at least as accurate as the best TSVD solution x_{k_0} . This problem has been intensively studied but has had no definitive solutions. For Krylov iterative solvers, their regularizing effects critically rely on how well the underlying k -dimensional Krylov subspace captures the k dominant right singular vectors of A [8, 10]. The richer information the Krylov subspace contains on the k dominant right singular vectors, the less possible it is that the resulting projected problem has a small singular value. That is, the solvers capture the large SVD components of A more effectively, and thus have better regularizing effects.

To precisely measure the regularizing effects, we introduce the term of *full* or *partial* regularization. If a pure iterative solver itself computes a best possible regularized solution at semi-convergence, it is said to have the full regularization; in this case, no additional regularization is necessary. Otherwise, it is said to have the partial regularization; in this case, a sophisticated hybrid variant is needed that combines the solver with some additional regularization in order to improve the accuracy of the regularized solution by the iterative solver at semi-convergence [8, 10]. It appears that the regularizing effects are closely related to the degree of ill-posedness of the problem. To this end, we introduce the following definition of the degree of ill-posedness, which follows Hofmann's book [12] and has been commonly used in the literature, e.g., [8, 10]: If there exists a positive real number α such that the singular values satisfy $\sigma_j = \mathcal{O}(j^{-\alpha})$, the problem is termed as mildly or moderately ill-posed if $\alpha \leq 1$ or $\alpha > 1$; if $\sigma_j = \mathcal{O}(e^{-\alpha j})$ with $\alpha > 0$ considerably, $j = 1, 2, \dots, n$, the problem is termed severely ill-posed. Clearly, the singular values σ_j of a severely ill-posed problem decay exponentially at the same rate $e^{-\alpha}$, while those of a moderately or mildly ill-posed problem decay more and more slowly at the decreasing rate $\left(\frac{j}{j+1}\right)^\alpha$ approaching one with increasing j , which, for the same j , is smaller for the moderately ill-posed problem than it for the mildly ill-posed problem.

For A symmetric, its k -dimensional dominant eigenspace is identical to the k -dimensional dominant left and right singular subspaces. In this case, MINRES and its variant MR-II are natural alternatives to LSQR and CGLS [5, 7, 10]. MR-II was originally designed for solving singular and inconsistent linear systems, and it uses the starting vector Ab and restricts the resulting Krylov subspace to the range of A . Thus, the iterates are orthogonal to the null space of A , and MR-II computes the minimum 2-norm least squares solution [3, 5]. For (1.1), we are not interested in such

solution but a regularized solution that is close to x_{true} as much as possible. MINRES and MR-II have been shown to have regularizing effects and exhibit semi-convergence [11, 14, 16], and MR-II usually provides better regularized solutions than MINRES. Intuitively, this is because the noise e in the initial Krylov vector Ab is filtered by multiplication with A [6, 14]. Different implementations associated with MR-II have been studied [5, 17].

In this paper, we first prove that the MINRES iterates are filtered SVD solutions, similar to the form of the LSQR iterates. Based on this result, we show why MINRES, in general, has only the partial regularization, independent of the degree of ill-posedness of (1.1). As a result, a hybrid MINRES that combines MINRES with a regularization method applied to the lower dimensional projected problems should be used to compute a best possible regularized solution; see [10, Section 6.4] for details. Afterwards, we take a closer look at the regularization of MINRES and MR-II in more detail, which, from a new perspective, shows that MINRES has only the partial regularization. We prove that, though MR-II has globally better regularizing effects than MINRES, the k th MINRES iterate is always more accurate than the $(k - 1)$ th MR-II iterate until the semi-convergence of MINRES. In a manner different from those used in [14, 16], we then analyze the regularizing effects of MR-II and draw some definitive conclusions. We establish bounds for the 2-norm distance between the underlying k -dimensional Krylov subspace and the k -dimensional dominant eigenspace. The bounds indicate that the Krylov subspace better captures the k -dimensional dominant eigenspace for severely and moderately ill-posed problems than for mildly ill-posed problems. As a consequence, MR-II has better regularizing effects for the first two kinds of problems than for the third kind, for which MR-II has only the partial regularization. We then use the results to derive an estimate for the accuracy of the rank k approximation generated by MR-II to A . Finally, we derive estimates for the entries generated by the symmetric Lanczos process that MR-II is based on, and show how fast they decay.

The paper is organized as follows. In Section 2, we describe MINRES and MR-II. In Section 3, we prove that the MINRES iterates are filtered SVD solutions, followed by an analysis on the regularizing effects of MINRES. In Section 4, we compare the regularizing effects of MINRES and MR-II, and shed light on some new features of them. In Section 5, we present our theoretical results on MR-II with a detailed analysis. In Section 6, we numerically confirm our theory that MINRES has only the partial regularization for a general ill-posed problem and its hybrid variant is needed. Also, we experimentally illustrate that MR-II has the full regularization for severely and moderately ill-posed problems, which is stronger than our theory, and it has the partial regularization for mildly ill-posed problems. We also compare MR-II with LSQR, demonstrating that MR-II is as effective as and at least twice as efficient as LSQR. We conclude the paper in Section 7.

Throughout the paper, we denote by $\mathcal{K}_k(C, w) = \text{span}\{w, Cw, \dots, C^{k-1}w\}$ the k -dimensional Krylov subspace generated by the matrix C and the vector w , by $\|\cdot\|$ and $\|\cdot\|_F$ the 2-norm of a matrix or vector and the Frobenius norm of a matrix, respectively, and by I the identity matrix with order clear from the context.

2 MINRES and MR-II

MINRES [20] is based on the symmetric Lanczos process that constructs an orthonormal basis of the Krylov subspace $\mathcal{K}_k(A, b)$. Let $\bar{q}_1 = b/\|b\|$. The k -step symmetric Lanczos process can be written in the matrix form

$$A\bar{Q}_k = \bar{Q}_{k+1}\bar{T}_k,$$

where $\bar{Q}_{k+1} = (\bar{q}_1, \bar{q}_2, \dots, \bar{q}_{k+1})$ has orthonormal columns which form $\mathcal{K}_k(A, b)$, and $\bar{T}_k \in \mathbb{R}^{(k+1) \times k}$ is a tridiagonal matrix with its leading $k \times k$ submatrix symmetric.

At iteration k , MINRES solves $\|b - A\bar{x}^{(k)}\| = \min_{x \in \mathcal{K}_k(A, b)} \|b - Ax\|$ for the iterate $\bar{x}^{(k)} = \bar{Q}_k \bar{y}^{(k)}$ with

$$\bar{y}^{(k)} = \arg \min_{y \in \mathbb{R}^k} \|b\|e_1 - \bar{T}_k y\|, \quad (2.1)$$

where e_1 is the first canonical vector of dimension $k+1$. For our analysis purpose, it is important to write

$$\bar{x}^{(k)} = \bar{Q}_k \bar{T}_k^\dagger \bar{Q}_{k+1}^T b, \quad (2.2)$$

which is the minimum 2-norm least squares solution of the perturbed problem that replace A in (1.1) by its rank k approximation $\bar{Q}_{k+1} \bar{T}_k \bar{Q}_k^T$.

MR-II [5] is a variant of MINRES applied to $\mathcal{K}_k(A, Ab)$ which excludes the noisy b . The method is based on the k -step symmetric Lanczos process

$$AQ_k = Q_{k+1} T_k, \quad (2.3)$$

where $Q_{k+1} = (q_1, q_2, \dots, q_{k+1})$ has orthonormal columns with $q_1 = Ab/\|Ab\|$, $T_k \in \mathbb{R}^{(k+1) \times k}$ is a tridiagonal matrix with the diagonals α_i , the subdiagonals $\beta_i > 0$, $i = 1, 2, \dots, k$ and the superdiagonals $\beta_i > 0$, $i = 1, 2, \dots, k-1$, and the first k rows of T_k is symmetric. The columns of Q_k form an orthonormal basis of $\mathcal{K}_k(A, Ab)$. Mathematically, since the eigenvalues of A are simple and Ab has nonzero components in the directions of all the eigenvectors v_i of A , the Lanczos process can be run to n steps without breakdown, i.e., $\beta_k > 0$, $k = 1, 2, \dots, n-1$ and $\beta_n = 0$.

At iteration k , MR-II solves $\|b - Ax^{(k)}\| = \min_{x \in \mathcal{K}_k(A, Ab)} \|b - Ax\|$ for the iterate $x^{(k)} = Q_k y^{(k)}$ with

$$y^{(k)} = \arg \min_{y \in \mathbb{R}^k} \|b - Q_{k+1} T_k y\|. \quad (2.4)$$

Similar to (2.2), we have the expression

$$x^{(k)} = Q_k T_k^\dagger Q_{k+1}^T b, \quad (2.5)$$

which is the minimum 2-norm least squares solution of the perturbed problem that replace A in (1.1) by its rank k approximation $Q_{k+1} T_k Q_k^T$.

The significance of (2.2) and (2.5) lies that the MINRES and MR-II iterates are the minimum 2-norm least squares solutions of the perturbed problems that replace A in (1.1) by its rank k approximations $\bar{Q}_{k+1} \bar{T}_k \bar{Q}_k^T$ and $Q_{k+1} T_k Q_k^T$, respectively, whose k nonzero singular values are just those of \bar{T}_k and T_k , respectively. If the singular values of \bar{T}_k or T_k approximate the k large singular values of A in natural order for $k = 1, 2, \dots, k_0$, then $\bar{Q}_{k+1} \bar{T}_k \bar{Q}_k^T$ and $Q_{k+1} T_k Q_k^T$ are near best rank k approximations to A with accuracy similar to that of the best rank k approximation A_k . If this is the case, MINRES and MR-II must have the full regularization, and $\bar{x}^{(k_0)}$ and $x^{(k_0)}$ are best possible regularized solutions and are as accurate as the best TSVD regularized solution x_{k_0} .

3 The regularizing effects of MINRES

Similar to the CGLS and LSQR iterates [8, p. 146], we can establish the following result on the MINRES iterates.

Theorem 3.1. *For MINRES to solve (1.1) with the starting vector $\bar{q}_1 = b/\|b\|$, the k th iterate $\bar{x}^{(k)}$ has the form*

$$\bar{x}^{(k)} = \sum_{i=1}^n f_i^{(k)} \frac{v_i^T b}{\lambda_i} v_i, \quad (3.1)$$

where the filters $f_i^{(k)} = 1 - \prod_{j=1}^k \frac{\theta_j^{(k)} - \lambda_i}{\theta_j^{(k)}}$, $i = 1, 2, \dots, n$ with $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n| > 0$, and $\theta_j^{(k)}$, $j = 1, 2, \dots, k$, are the harmonic Ritz values of A with respect to $\mathcal{K}_k(A, b)$ and labeled as $|\theta_1^{(k)}| > |\theta_2^{(k)}| > \dots > |\theta_k^{(k)}| > 0$.

Proof. From [18], the residual $\bar{r}^{(k)} = b - A\bar{x}^{(k)}$ of the MINRES iterate $\bar{x}^{(k)}$ can be written as

$$\bar{r}^{(k)} = \chi_k(A)b, \quad (3.2)$$

where the residual polynomial $\chi_k(t)$ has the form

$$\chi_k(t) = \prod_{j=1}^k \frac{\theta_j^{(k)} - t}{\theta_j^{(k)}}$$

with the $\theta_j^{(k)}$ the harmonic Ritz values of A with respect to $\mathcal{K}_k(A, b)$. From (3.2), we get

$$\bar{x}^{(k)} = (I - \chi_k(A))A^{-1}b.$$

Substituting $A = V\Lambda V^T$ into the above gives

$$\bar{x}^{(k)} = \sum_{i=1}^n f_i^{(k)} \frac{v_i^T b}{\lambda_i} v_i,$$

where

$$f_i^{(k)} = 1 - \prod_{j=1}^k \frac{\theta_j^{(k)} - \lambda_i}{\theta_j^{(k)}}, \quad i = 1, 2, \dots, n. \quad \square$$

Relation (3.1) shows that the MINRES iterate $\bar{x}^{(k)}$ has a filtered SVD expansion. For a general symmetric A , the harmonic Ritz values have an attractive feature: they usually favor extreme eigenvalues of A , provided that a Krylov subspace contains substantial information on all the eigenvectors v_i [18]. In our current context, if at least a small harmonic Ritz value in magnitude starts to appear for some $k \leq k_0$, i.e., $|\theta_k^{(k)}| \leq |\lambda_{k_0+1}|$, the corresponding filter factors $f_i^{(k)}$, $i = k+1, \dots, n$, are not small, meaning that $\bar{x}^{(k)}$ is already deteriorated. On the other hand, if no small harmonic Ritz value in magnitude appears before $k \leq k_0$, the $\bar{x}^{(k)}$ are expected to become better approximations to x_{true} until $k = k_0$. Unfortunately, since $\mathcal{K}_k(A, b)$ includes the noisy $b = \hat{b} + e$, which contains non-negligible components of v_i corresponding to small eigenvalues λ_i , it is generally possible that a small harmonic Ritz value can appear for $k \leq k_0$. This demonstrates that, in general, MINRES only has the partial regularization and cannot obtain a best possible regularized solution.

4 Regularization relationships between MINRES and MR-II

It was known a long time ago that MR-II has better regularizing effects than MINRES, that is, MR-II obtains a more accurate regularized solution than MINRES does [5]. Such phenomenon is simply due to the fact that $\mathcal{K}_k(A, b)$ for MINRES includes the noisy b and $\mathcal{K}_k(A, Ab)$ for MR-II contains less information on v_i corresponding to small eigenvalues in magnitude since the noise e in the starting vector Ab is filtered by multiplication with A . Previously, we have given an analysis on the regularizing effects of MINRES and shown that a hybrid MINRES is generally needed for

an ill-posed problem, independent of the degree of ill-posedness of (1.1). Next we shed more light on the regularization of MINRES, compare it with MR-II, and reveal some new features of them.

To simplify our discussions, without loss of generality, we can well assume that for a standard nonsingular linear system, the smaller residual, the more accurate the approximate solution is. Given the residual minimization property of MINRES and MR-II, one might be confused that, since $\mathcal{K}_{k-1}(A, Ab) \subset \mathcal{K}_k(A, b)$, the k th MINRES iterate $\bar{x}^{(k)}$ should be at least as accurate as the $(k-1)$ th MR-II iterate $x^{(k-1)}$. This is true for solving the standard linear system where the right-hand side is supposed to be *exact*, but it is nontrivial and depends for solving an ill-posed problem, for which the b is noisy and we are concerned with regularized approximations to the true solution x_{true} other than the naive solution x_{naive} . Our previous analysis has shown that a small harmonic Ritz value $|\theta_k^{(k)}| \leq \sigma_{k_0+1} = |\lambda_{k_0+1}|$ generally appears for MINRES before some iteration $k \leq k_0$, causing that MINRES has only the partial regularization. On the other hand, however, note that the regularized solutions $\bar{x}^{(k)}$ by MINRES converge to x_{true} until the semi-convergence of MINRES. As a result, because $\mathcal{K}_{k-1}(A, Ab) \subset \mathcal{K}_k(A, b)$, the k th MINRES iterate $\bar{x}^{(k)}$ is more accurate than the $(k-1)$ th MR-II iterate $x^{(k-1)}$ *only until* the semi-convergence of MINRES.

We can also explain the partial regularization of MINRES in terms of the rank k approximation $\bar{Q}_{k+1}\bar{T}_k\bar{Q}_k^T$ to A as follows: Since the k -dimensional dominant eigenspace of A is identical to its k -dimensional dominant left and right singular subspaces, $\mathcal{K}_k(A, b)$ contains substantial information on all the v_i . As a result, it is generally possible that the projected matrix \bar{T}_k has a singular value smaller than σ_{k_0+1} for some $k \leq k_0$. This means that $\bar{Q}_{k+1}\bar{T}_k\bar{Q}_k^T$ is a poor rank k approximation to A , causing, from (2.1), that $\|\bar{x}^{(k)}\| = \|\bar{Q}_k\bar{y}^{(k)}\| = \|b\|\|\bar{T}_k^\dagger e_1\|$ is generally large, i.e., $\bar{x}^{(k)}$ is already deteriorated. Conversely, if no singular value of \bar{T}_k is smaller than σ_{k_0+1} and the semi-convergence of MINRES does not yet occur, the MINRES iterate $\bar{x}^{(k)}$ should be at least as accurate as the MR-II iterate $x^{(k-1)}$ because of $\mathcal{K}_{k-1}(A, Ab) \subset \mathcal{K}_k(A, b)$.

In summary, we need to use a hybrid MINRES with the TSVD method or the standard-form Tikhonov regularization applied to the projected problem in (2.1) to expand the Krylov subspace until it contains all the k_0 dominant spectral components and a best regularized solution is found, in which the additional regularization aims to remove the effects of small singular values of \bar{T}_{k+1} , similar to the hybrid LSQR see [10, Section 6.4].

5 Regularizing effects of MR-II

Before proceeding, we point out that, unlike (3.1) for the MINRES iterates $\bar{x}^{(k)}$, we have found that the MR-II iterates $x^{(k)}$ do not have filtered SVD expansions of similar form. Even so, we can establish a number of other results that help to better understand the regularization of MR-II. We first investigate a fundamental problem: how well does the underlying subspace $\mathcal{K}_k(A, Ab)$ capture the k dimensional dominant eigenspace of A ? This problem is of basic importance because it critically affects the accuracy of $Q_{k+1}T_kQ_k^T$ as a rank k approximation to A .

In terms of the definition of canonical angles $\Theta(\mathcal{X}, \mathcal{Y})$ between the two subspaces \mathcal{X} and \mathcal{Y} of the same dimension [22, p. 250], we present the following result.

Theorem 5.1. *Let $A = V\Omega\Sigma V^T = V\Lambda V^T$ be defined as (1.2), and assume that the singular values of A are $\sigma_j = |\lambda_j| = \mathcal{O}(e^{-\alpha j})$ with $\alpha > 0$. Let $\mathcal{V}_k = \text{span}\{V_k\}$ be the k -dimensional dominant spectral subspace spanned by the columns of $V_k = (v_1, v_2, \dots, v_k)$, and $\mathcal{V}_k^s = \mathcal{K}_k(A, Ab)$. Then*

$$\|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\| = \frac{\|\Delta_k\|}{\sqrt{1 + \|\Delta_k\|^2}} \quad (5.1)$$

with the $(n - k) \times k$ matrix Δ_k to be defined by (5.3) and

$$\|\Delta_k\|_F \leq \frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{\max_{j=k+1}^n |v_j^T b|}{\min_{j=1}^k |v_j^T b|} \sqrt{k(n-k)} (1 + \mathcal{O}(e^{-\alpha})), \quad k = 1, 2, \dots, n-1. \quad (5.2)$$

Proof. Note that $\mathcal{K}_k(\Lambda, \Lambda V^T b)$ is spanned by the columns of the $n \times k$ matrix DB_k with

$$D = \text{diag}(\lambda_i v_i^T b), \quad B_k = \begin{pmatrix} 1 & \lambda_1 & \dots & \lambda_1^{k-1} \\ 1 & \lambda_2 & \dots & \lambda_2^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \dots & \lambda_n^{k-1} \end{pmatrix}.$$

Partition D and B_k as follows:

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}, \quad B_k = \begin{pmatrix} B_{k1} \\ B_{k2} \end{pmatrix},$$

where $D_1, B_{k1} \in \mathbb{R}^{k \times k}$. Since B_{k1} is a Vandermonde matrix with λ_j distinct for $1 \leq j \leq k$, it is nonsingular. Noting $\mathcal{K}_k(A, Ab) = V\mathcal{K}_k(\Lambda, \Lambda V^T b)$, we have

$$\mathcal{K}_k(A, Ab) = \text{span}\{VDB_k\} = \text{span}\left\{V \begin{pmatrix} D_1 B_{k1} \\ D_2 B_{k2} \end{pmatrix}\right\} = \text{span}\left\{V \begin{pmatrix} I \\ \Delta_k \end{pmatrix}\right\}$$

with

$$\Delta_k = D_2 B_{k2} B_{k1}^{-1} D_1^{-1}. \quad (5.3)$$

Define $Z_k = V \begin{pmatrix} I \\ \Delta_k \end{pmatrix}$. Then $Z_k^T Z_k = I + \Delta_k^T \Delta_k$, and the columns of $Z_k (Z_k^T Z_k)^{-\frac{1}{2}}$ form an orthonormal basis of \mathcal{V}_k^s .

Write $V = (V_k, V_k^\perp)$. By definition, we obtain

$$\begin{aligned} \|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\| &= \left\| (V_k^\perp)^T Z_k (Z_k^T Z_k)^{-\frac{1}{2}} \right\| \\ &= \left\| (V_k^\perp)^T V \begin{pmatrix} I \\ \Delta_k \end{pmatrix} (I + \Delta_k^T \Delta_k)^{-\frac{1}{2}} \right\| \\ &= \|\Delta_k (I + \Delta_k^T \Delta_k)^{-1/2}\| \\ &= \frac{\|\Delta_k\|}{\sqrt{1 + \|\Delta_k\|^2}}, \end{aligned}$$

which proves (5.1).

We next estimate $\|\Delta_k\|$ and establish upper bound for the right-hand side of (5.1). We have

$$\begin{aligned} \|\Delta_k\| &\leq \|\Delta\|_F = \|D_2 B_{k2} B_{k1}^{-1} D_1^{-1}\|_F \leq \|D_2\| \|B_{k2} B_{k1}^{-1}\|_F \|D_1^{-1}\| \\ &= \frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{\max_{j=k+1}^n |v_j^T b|}{\min_{j=1}^k |v_j^T b|} \|B_{k2} B_{k1}^{-1}\|_F. \end{aligned} \quad (5.4)$$

We now estimate $\|B_{k2} B_{k1}^{-1}\|_F$. It is easily justified that the i th column of B_{k1}^{-1} consists of the coefficients of the Lagrange polynomial

$$L_i^{(k)}(\lambda) = \prod_{j=1, j \neq i}^k \frac{\lambda - \lambda_j}{\lambda_i - \lambda_j}$$

that interpolates the elements of the i th canonical basis vector $e_i^{(k)} \in \mathbb{R}^k$ at the abscissas $\lambda_1, \dots, \lambda_k$. Consequently, the i th column of $B_{k2}B_{k1}^{-1}$ is

$$B_{k2}B_{k1}^{-1}e_i^{(k)} = \left(L_i^{(k)}(\lambda_{k+1}), \dots, L_i^{(k)}(\lambda_n) \right)^T,$$

from which we obtain

$$B_{k2}B_{k1}^{-1} = \begin{pmatrix} L_1^{(k)}(\lambda_{k+1}) & L_2^{(k)}(\lambda_{k+1}) & \dots & L_k^{(k)}(\lambda_{k+1}) \\ L_1^{(k)}(\lambda_{k+2}) & L_2^{(k)}(\lambda_{k+2}) & \dots & L_k^{(k)}(\lambda_{k+2}) \\ \vdots & \vdots & \ddots & \vdots \\ L_1^{(k)}(\lambda_n) & L_2^{(k)}(\lambda_n) & \dots & L_k^{(k)}(\lambda_n) \end{pmatrix}. \quad (5.5)$$

For a fixed λ satisfying $|\lambda| \leq |\lambda_{k+1}|$, let $i_0 = \arg \max_{i=1,2,\dots,k} |L_i^{(k)}(\lambda)|$. Then we have

$$|L_{i_0}^{(k)}(\lambda)| = \prod_{j=1, j \neq i_0}^k \left| \frac{\lambda - \lambda_j}{\lambda_{i_0} - \lambda_j} \right| \leq \prod_{j=1, j \neq i_0}^k \left| \frac{|\lambda_j - \lambda|}{|\lambda_j| - |\lambda_{i_0}|} \right| \leq \prod_{j=1, j \neq i_0}^k \left| \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_j| - |\lambda_{i_0}|} \right|. \quad (5.6)$$

Therefore, for $i = 1, 2, \dots, k$ and $|\lambda| \leq |\lambda_{k+1}|$, making use of Taylor series expansions, we get

$$\begin{aligned} |L_i^{(k)}(\lambda)| &\leq \prod_{j=1, j \neq i_0}^k \left| \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_j| - |\lambda_{i_0}|} \right| = \prod_{j=1}^{i_0-1} \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_j| - |\lambda_{i_0}|} \cdot \prod_{j=i_0+1}^k \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_{i_0}| - |\lambda_j|} \\ &= \prod_{j=1}^{i_0-1} \frac{1 + \mathcal{O}(e^{-(k-j+1)\alpha})}{1 - \mathcal{O}(e^{-(i_0-j)\alpha})} \cdot \prod_{j=i_0+1}^k \frac{\mathcal{O}(e^{-(k-j+1)\alpha}) + 1}{\mathcal{O}(e^{(j-i_0)\alpha}) - 1} \\ &= \frac{\prod_{j=1}^{i_0-1} (1 + \mathcal{O}(e^{-(k-j+1)\alpha}))}{1 + \mathcal{O}(e^{-(k-i_0+1)\alpha})} \prod_{j=1}^{i_0-1} \frac{1}{1 - \mathcal{O}(e^{-(i_0-j)\alpha})} \prod_{j=i_0+1}^k \frac{1}{\mathcal{O}(e^{(j-i_0)\alpha}) - 1} \\ &= \frac{\prod_{j=1}^k (1 + \mathcal{O}(e^{-(k-j+1)\alpha}))}{(1 + \mathcal{O}(e^{-(k-i_0+1)\alpha}))} \prod_{j=1}^{i_0-1} \frac{1}{1 - \mathcal{O}(e^{-(i_0-j)\alpha})} \\ &\quad \cdot \prod_{j=i_0+1}^k \frac{1}{1 - \mathcal{O}(e^{-(j-i_0)\alpha})} \frac{1}{\prod_{j=i_0+1}^k \mathcal{O}(e^{(j-i_0)\alpha})} \\ &= \frac{\left(1 + \sum_{j=1}^{k+1} \mathcal{O}(e^{-(k-j+1)\alpha}) \right)}{(1 + \mathcal{O}(e^{-(k-i_0+1)\alpha}))} \frac{\left(1 + \sum_{j=1}^{i_0} \mathcal{O}(e^{-j\alpha}) \right)}{\prod_{j=i_0+1}^k \mathcal{O}(e^{(j-i_0)\alpha})} \frac{\left(1 + \sum_{j=1}^{k-i_0+1} \mathcal{O}(e^{-j\alpha}) \right)}{\prod_{j=i_0+1}^k \mathcal{O}(e^{(j-i_0)\alpha})} \end{aligned} \quad (5.7)$$

by absorbing those higher order terms into $\mathcal{O}(\cdot)$. Note that in the above numerator we have

$$1 + \sum_{j=1}^{k+1} \mathcal{O}(e^{-(k-j+1)\alpha}) = 1 + \mathcal{O}\left(\sum_{j=1}^{k+1} e^{-(k-j+1)\alpha}\right) = 1 + \mathcal{O}\left(\frac{e^{-\alpha}}{1 - e^{-\alpha}}(1 - e^{-(k+1)\alpha})\right),$$

$$1 + \sum_{j=1}^{i_0} \mathcal{O}(e^{-j\alpha}) = 1 + \mathcal{O}\left(\sum_{j=1}^{i_0} e^{-j\alpha}\right) = 1 + \mathcal{O}\left(\frac{e^{-\alpha}}{1 - e^{-\alpha}}(1 - e^{-i_0\alpha})\right),$$

and

$$1 + \sum_{j=1}^{k-i_0+1} \mathcal{O}(e^{-j\alpha}) = 1 + \mathcal{O}\left(\sum_{j=1}^{k-i_0+1} e^{-j\alpha}\right) = 1 + \mathcal{O}\left(\frac{e^{-\alpha}}{1 - e^{-\alpha}}(1 - e^{-(k-i_0+1)\alpha})\right).$$

It is easy to check that for any $1 \leq i_0 \leq k$ the product of the above three terms is no more than

$$1 + \mathcal{O}\left(\frac{3e^{-\alpha}}{1 - e^{-\alpha}}\right) + \mathcal{O}\left(\left(\frac{e^{-\alpha}}{1 - e^{-\alpha}}\right)^2\right) = 1 + \mathcal{O}(e^{-\alpha}).$$

By definition, the factor $\prod_{j=i_0+1}^k \mathcal{O}(e^{(j-i_0)\alpha}) = \prod_{j=i_0+1}^k \frac{|\lambda_{i_0}|}{|\lambda_j|}$ in the denominator of (5.7), which is exactly one when $i_0 = k$, and it is bigger than one when $i_0 < k$; the other factor $1 + \mathcal{O}(e^{-(k-i_0+1)\alpha})$ is between $1 + \mathcal{O}(e^{-k\alpha})$ and $1 + \mathcal{O}(e^{-\alpha})$. Therefore, for any k and $|\lambda| \leq |\lambda_{k+1}|$, we have

$$|L_k^{(k)}(\lambda)| = 1 + \mathcal{O}(e^{-\alpha}), \quad (5.8)$$

$$|L_{i_0}^{(k)}(\lambda)| = \max_{i=1,2,\dots,k} |L_i^{(k)}(\lambda)| = 1 + \mathcal{O}(e^{-\alpha}). \quad (5.9)$$

From this estimate and (5.5) it follows that

$$\|B_{k2}B_{k1}^{-1}\|_F \leq \sqrt{k(n-k)} (1 + \mathcal{O}(e^{-\alpha})). \quad (5.10)$$

As a result, for $k = 1, 2, \dots, n-1$, from (5.4) we have

$$\|\Delta\|_F \leq \frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{\max_{j=k+1}^n |v_j^T b|}{\min_{j=1}^k |v_j^T b|} \sqrt{k(n-k)} (1 + \mathcal{O}(e^{-\alpha})). \quad \square$$

Remark 5.1 Trivially, we have

$$\|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\| \leq 1.$$

But in our context it is impossible to have $\|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\| = 1$ since Δ_k is not a zero matrix. We have seen from the proof that the factor $\frac{|\lambda_{k+1}|}{|\lambda_k|} \frac{\max_{j=k+1}^n |v_j^T b|}{\min_{j=1}^k |v_j^T b|}$ in it is intrinsic and unavoidable in (5.2). But the factor $\sqrt{k(n-k)}$ in (5.2) is an overestimate and can certainly be reduced. The reason is that (5.10) is an overestimate since $|L_i^{(k)}(\lambda_j)|$ for i not near to k is considerably smaller than $|L_{i_0}^{(k)}(\lambda_j)|$, $j = k+1, \dots, n$ but we replace all them by the maximum $|L_{i_0}^{(k)}(\lambda_j)| = 1 + \mathcal{O}(e^{-\alpha})$. In fact, our derivation before (5.8) and (5.9) when replacing i_0 by i clearly illustrates that the smaller i is, the smaller $|L_i^{(k)}(\lambda_j)|$ than $|L_k^{(k)}(\lambda_j)|$, $j = k+1, \dots, n$.

Recall the discrete Picard condition (1.4). Then the coefficients

$$c_k = \frac{\max_{j=k+1}^n |v_j^T b|}{\min_{j=1}^k |v_j^T b|} = \frac{\max_{j=k+1}^n (|v_j^T \hat{b} + v_j^T e|)}{\min_{j=1}^k (|v_j^T \hat{b} + v_j^T e|)} \approx \frac{|\lambda_{k+1}|^{1+\beta} + |v_{k+1}^T e|}{|\lambda_k|^{1+\beta} + |v_k^T e|}. \quad (5.11)$$

We see that, the larger β is, the smaller $c_k \approx \frac{|\lambda_{k+1}|^{1+\beta}}{|\lambda_k|^{1+\beta}}$, which is a constant for $k \leq k_0$, and thus the better \mathcal{V}_k^s captures \mathcal{V}_k . For $k > k_0$, since all the $|v_k^T b| \approx |v_k^T e|$ are roughly the same, we have $c_k \approx 1$, meaning that \mathcal{V}_k^s may not capture \mathcal{V}_k so well after iteration k_0 .

Remark 5.2 The theorem can be extended to moderately ill-posed problems with the singular values $\sigma_j = \mathcal{O}(j^{-\alpha})$, $\alpha > 1$ considerably and k not big, where the factor $1 + \mathcal{O}(e^{-\alpha})$ in (5.2) is replaced by a bigger $\mathcal{O}(1)$. Let us look into why it is so. Recall that, by definition, $|L_{i_0}^{(k)}(\lambda)| \geq |L_k^{(k)}(\lambda)|$ for $|\lambda| \leq |\lambda_{k+1}|$. Using a similar proof to that of Theorem 5.1 and the first order Taylor expansion, we can roughly estimate $|L_k^{(k)}(\lambda)|$ as follows:

$$\begin{aligned} |L_{i_0}^{(k)}(\lambda)| &\approx |L_k^{(k)}(\lambda)| \leq \prod_{j=1}^{k-1} \left| \frac{|\lambda_j| + |\lambda_{k+1}|}{|\lambda_j| - |\lambda_k|} \right| \\ &= \prod_{j=1}^{k-1} \frac{1 + \mathcal{O}\left(\left(\frac{j}{k+1}\right)^\alpha\right)}{1 - \mathcal{O}\left(\left(\frac{j}{k}\right)^\alpha\right)} \\ &\approx \sum_{j=1}^{k-1} \left(1 + \mathcal{O}\left(\left(\frac{j}{k+1}\right)^\alpha\right)\right) \cdot \sum_{j=1}^{k-1} \left(1 + \mathcal{O}\left(\left(\frac{j}{k}\right)^\alpha\right)\right) = \mathcal{O}(1). \end{aligned}$$

This estimate is not as accurate as that for severely ill-posed problems. More important is that it depends on k and increases slowly as k increases. The above estimate can be improved when A is symmetric definite:

$$\begin{aligned} |L_{i_0}^{(k)}(\lambda)| &\approx |L_k^{(k)}(\lambda)| = \prod_{j=1}^{k-1} \frac{|\lambda_j - \lambda_{k+1}|}{|\lambda_j - \lambda_k|} \leq \prod_{j=1}^{k-1} \frac{|\lambda_j|}{|\lambda_j - \lambda_k|} \\ &= \prod_{j=1}^{k-1} \frac{1}{1 - \mathcal{O}\left(\left(\frac{j}{k}\right)^\alpha\right)} \\ &\approx \sum_{j=1}^{k-1} \left(1 + \mathcal{O}\left(\left(\frac{j}{k}\right)^\alpha\right)\right) = \mathcal{O}(1), \end{aligned}$$

smaller than the previous one. The two estimates mean that \mathcal{V}_k^s may capture \mathcal{V}_k better for A symmetric definite than for A symmetric indefinite where there are both positive and negative ones among the first $k + 1$ large eigenvalues.

Remark 5.3 A combination of (5.1) and (5.2) and the above analysis indicate that \mathcal{V}_k^s captures \mathcal{V}_k better for severely ill-posed problems than for moderately ill-posed problems. There are two reasons for this. The first is that the factors $|\lambda_{k+1}/\lambda_k|$ are basically fixed constants for severely ill-posed problems as k increases, and they are smaller than the counterparts for moderately ill-posed problems unless the degree α of its ill-posedness is far bigger than one and k small. The second is that the factor $1 + \mathcal{O}(e^{-\alpha})$ is smaller for severely ill-posed problems than the factor $\mathcal{O}(1)$ for moderately ill-posed problems.

Remark 5.4 The situation is fundamentally different for mildly ill-posed problems: Firstly, we always have $|L_{i_0}^{(k)}(\lambda)| > 1$ substantially for $|\lambda| \leq |\lambda_{k+1}|$, $\alpha \leq 1$ and any k . Secondly, c_k defined by (5.11) is closer to one than that for moderately ill-posed problems for $k = 1, 2, \dots, k_0$. Thirdly, for the same noise level $\|e\|$ and β , from the discrete Picard condition (1.4) and the definition of k_0 we see that k_0 is bigger for a mildly ill-posed problem than that for a moderately ill-posed problem. All of them show that \mathcal{V}_k^s captures \mathcal{V}_k *considerably better* for severely and moderately ill-posed problems than for mildly ill-posed problems. In other words, our results imply that \mathcal{V}_k^s contains substantial information on the other $n - k$ eigenvectors for mildly ill-posed problems, causing that a small harmonic Ritz value generally appears for some $k \leq k_0$, especially when k_0 is not small.

Equivalently, the projected matrix T_k generated by MR-II generally has a small singular value for some $k \leq k_0$, such that the solution $x^{(k)}$ is deteriorated, as deduced from (2.5). As a result, we are certain that MR-II has better regularizing effects for severely and moderately ill-posed problems than for mildly ill-posed problems. Most importantly, by this property, since MR-II has *at most* the full regularization for severely and moderately ill-posed problems, we deduce and are thus sure that MR-II generally has only the partial regularization for mildly ill-posed problems.

We mention that, in comparison with the results, i.e., Theorem 2.1, in [13] on LSQR, we find that $\mathcal{K}_k(A, Ab)$ is as comparably effective as $\mathcal{K}_k(A^T A, A^T b)$, on which LSQR works, for capturing the k -dimensional dominant eigenspace.

Let us get more insight into the regularization of MR-II. Recall (2.5), where

$$x^{(k)} = (Q_{k+1} T_k Q_k^T)^\dagger b = Q_k T_k^\dagger Q_{k+1}^T b.$$

Define

$$\gamma_k = \|A - Q_{k+1} T_k Q_k^T\|, \quad (5.12)$$

which measures the quality or accuracy of the rank k approximation $Q_{k+1} T_k Q_k^T$ to A . This quantity is central and fundamental to understand the regularizing effects of MR-II and measures how the iterates $x^{(k)}$ by MR-II behave like the TSVD regularized solution $x_k = A_k^\dagger b$. Particularly, note that the best rank k_0 approximation A_{k_0} satisfies $\|A - A_{k_0}\| = \sigma_{k_0+1}$. Then if $\gamma_{k_0} \approx \sigma_{k_0+1}$ for σ_{k_0+1} reasonably small, $Q_{k_0+1} T_{k_0} Q_{k_0}^T$ is a near best rank k_0 approximation to A with approximate accuracy σ_{k_0+1} and has no small nonzero singular value. In this case, the regularized solution $x^{(k_0)}$ is close to the best TSVD regularized solution x_{k_0} , and MR-II has the full regularization. Otherwise, if $\gamma_{k_0} > \sigma_{k_0+1}$ considerably, then $Q_{k_0+1} T_{k_0} Q_{k_0}^T$ deviates from the best rank k approximation A_{k_0} considerably and $x^{(k_0)}$ is not close to x_{k_0} , meaning that MR-II has only the partial regularization.

Based on Theorem 5.1 and Remark 5.2, we can derive the following estimates for γ_k .

Theorem 5.2. *Assume that (1.1) is severely or moderately ill posed. Then*

$$|\lambda_{k+1}| \leq \gamma_k \leq |\lambda_{k+1}| + |\lambda_1| \|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\|. \quad (5.13)$$

Proof. Note that $Q_{k+1} T_k Q_k^T$ is of rank k . The lower bound in (5.13) is trivial since the best k approximation A_k to A satisfies $\|A - A_k\| = \sigma_{k+1} = |\lambda_{k+1}|$. We next prove the upper bound. From (2.3), we obtain

$$\|A - Q_{k+1} T_k Q_k^T\| = \|A - A Q_k Q_k^T\| = \|A(I - Q_k Q_k^T)\|. \quad (5.14)$$

From Theorem 5.1, it is known that $\mathcal{V}_k^s = \mathcal{K}_k(A, Ab) = \text{span}\{Q_k\}$. Let $V_k = (v_1, v_2, \dots, v_k)$ and $\Lambda_k = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$. Then by the definition of $\|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\|$ we obtain

$$\begin{aligned} \|A - A Q_k Q_k^T\| &= \|(A - V_k \Lambda_k V_k^T + V_k \Lambda_k V_k^T)(I - Q_k Q_k^T)\| \\ &\leq \|(A - V_k \Lambda_k V_k^T)(I - Q_k Q_k^T)\| + \|V_k \Lambda_k V_k^T(I - Q_k Q_k^T)\| \\ &\leq |\lambda_{k+1}| + \|\Lambda_k\| \|V_k^T(I - Q_k Q_k^T)\| \\ &= |\lambda_{k+1}| + |\lambda_1| \|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\|. \quad \square \end{aligned}$$

Our later numerical experiments will indicate that $\gamma_k \approx \sigma_{k+1} = |\lambda_{k+1}|$ for severely and moderately ill-posed problems, illustrating that $Q_{k+1} T_k Q_k^T$ is a near best rank k approximation to A with the approximate accuracy σ_{k+1} . Particularly, since $\gamma_{k_0} \approx \sigma_{k_0+1}$, the MR-II iterate $x^{(k_0)} = Q_{k_0} T_{k_0}^\dagger Q_{k_0+1}^T b$ is close to the TSVD solution x_{k_0} provided that σ_{k_0+1} . Furthermore, we will

find that the error $\|x^{(k_0)} - x_{true}\|$ of MR-II iterate $x^{(k_0)}$ is as small as the error $\|x_{k_0} - x_{true}\|$ of the best TSVD solution x_{k_0} . This indicates that MR-II has the full regularization. Experimentally, for severely and moderately ill-posed problems, the observations $\gamma_k \approx \sigma_{k+1}$ appear to be general and thus should have strong theoretical supports. Our upper bound in (5.13) appears to be a considerable overestimate.

Recall that α_i and β_i , $i = 1, 2, \dots, k$ denote the diagonals and subdiagonals of T_k defined by (2.3), respectively. We next establish some interesting and intimate relationships between them and γ_k , showing how fast α_k and β_k decay.

Theorem 5.3. *For $k = 1, 2, \dots, n - 2$ we have*

$$\beta_{k+1} \leq \gamma_k, \quad (5.15)$$

$$|\alpha_{k+2}| \leq \gamma_k. \quad (5.16)$$

Proof. Since the Lanczos process can be run to completion, we have

$$Q_n^T A Q_n = \hat{T}_n,$$

where $Q_n \in \mathbb{R}^{n \times n}$ is orthogonal, and

$$\hat{T}_n = \begin{pmatrix} \alpha_1 & \beta_1 & & & \\ \beta_1 & \alpha_2 & \beta_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & \ddots & \beta_{n-1} \\ & & & \beta_{n-1} & \alpha_n \end{pmatrix} \quad (5.17)$$

is symmetric tridiagonal. Thus, from (2.3) we have

$$\begin{aligned} \gamma_k &= \|A - Q_{k+1} T_k Q_k^T\| = \|Q_n^T (A - Q_{k+1} T_k Q_k^T) Q_n\| \\ &= \left\| \hat{T}_n - \begin{pmatrix} I & \\ \mathbf{0} & \end{pmatrix} T_k \begin{pmatrix} I & \mathbf{0} \end{pmatrix} \right\| = \|G_k\|, \end{aligned}$$

where

$$G_k = \begin{pmatrix} \beta_{k+1} & \alpha_{k+2} & \beta_{k+2} & & \\ & \beta_{k+2} & \alpha_{k+3} & \beta_{k+4} & \\ & & \ddots & \ddots & \\ & & & \ddots & \beta_{n-1} \\ & & & \beta_{n-1} & \alpha_n \end{pmatrix} \in \mathbb{R}^{(n-k-1) \times (n-k)},$$

from which and $\beta_i > 0$, $i = 1, 2, \dots, n - 1$ it follows that

$$\beta_{k+1} = \|G_k e_1\| \leq \|G_k\| = \gamma_k$$

and

$$|\alpha_{k+2}| \leq \sqrt{\alpha_{k+2}^2 + \beta_{k+2}^2} = \|G_k e_2\| \leq \|G_k\| = \gamma_k$$

for $k = 1, 2, \dots, n - 2$. Therefore, (5.15) and (5.16) hold. \square

This theorem indicates that $|\alpha_{k+2}|$ and β_{k+1} decay at least as fast as γ_k . Moreover, based on the experimental observations that $\gamma_k \approx \sigma_{k+1}$ for severely and moderately ill-posed problems, the theorem illustrates that $|\alpha_{k+2}|$ and β_{k+1} decay as fast as σ_{k+1} , $k = 1, 2, \dots, n - 2$, for these two kinds of problems.

6 Numerical experiments

In this section, we report numerical experiments to illustrate the regularizing effects of MINRES and MR-II and make a number of comparisons. We justify our theory: (i) MINRES has only the partial regularization, independent of the degree of ill-posedness, and a hybrid MINRES is generally needed; (ii) the k th MINRES iterate $\bar{x}^{(k)}$ is always more accurate than the $(k-1)$ th MR-II iterate $x^{(k-1)}$ until the semi-convergence of MINRES; (iii) MR-II has only the partial regularization for mildly ill-posed problems, and a hybrid MR-II is needed. In the meantime, experimentally, we demonstrate a remarkable and attractive feature, stronger than our theory predicts: MR-II has the full regularization for severely and moderately ill-posed problems and its iterates at semi-convergence is as accurate as the best TSVD solutions for these two kinds of problems. We will use the function `lcurve` in [9] to depict the L-curves. In order to simulate exact arithmetic, the Lanczos process with reorthogonalization is used in MINRES and MR-II.

Table 1 lists test problems and their degree of ill-posedness, all of which are symmetric and arise from the discretization of the first kind Fredholm integral equations; see Hansen’s regularization toolbox [9] for details. For each problem except the 2D image deblurring problem ‘blur’, we use the corresponding code in [9] to generate a 1024×1024 matrix A , the true solution x_{true} and noise-free right-hand \hat{b} . In order to simulate the noisy data, we generate the white noise vector e whose entries are normally distributed with mean zero, such that the relative noise level $\varepsilon = \frac{\|e\|}{\|\hat{b}\|} = 10^{-2}, 10^{-3}, 10^{-4}$, respectively. To simulate exact arithmetic, the full reorthogonalization is used during the Lanczos process. We remind that, as far as ill-posed problem (1.1) is concerned, our primary goal consists in justifying the regularizing effects of iterative solvers, which are *unaffected by sizes* of ill-posed problems and only depends on the degree of ill-posedness. Therefore, for this purpose, as extensively done in the literature (see, e.g., [8, 10] and the references therein), it is enough to test not very large problems. Indeed, for n large, say, 1,0000, we have observed completely the same behavior as that for n not large, e.g., $n = 1024$ used in this paper except for the problem ‘blur’ with $n = 65,536$. A reason for using n not large is because such choice makes it practical to fully justify the regularization effects of LSQR by comparing it with the TSVD method, which suits only for small and/or medium sized problems for computational efficiency. All the computations are carried out using Matlab 7.8 with the machine precision $\epsilon_{mach} = 2.22 \times 10^{-16}$ under the Microsoft Windows 7 64-bit system.

Table 1: The description of test problems.

Problem	Description	Ill-posedness
shaw	One-dimensional image restoration model	severe
foxgood	Severely ill-posed test problem	severe
gravity	One-dimensional gravity surveying problem	severe
phillips	phillips’ ”famous” test problem	moderate
deriv2	Computation of second derivative	mild
blur	2D Image deblurring test problem	mild/moderate

6.1 A comparison of the regularizing effects of MR-II and MINRES

We now compare MINRES and MR-II and justify our theory: (i) the MR-II iterate is always more accurate than the MINRES iterate at their respective semi-convergence, meaning that MINRES

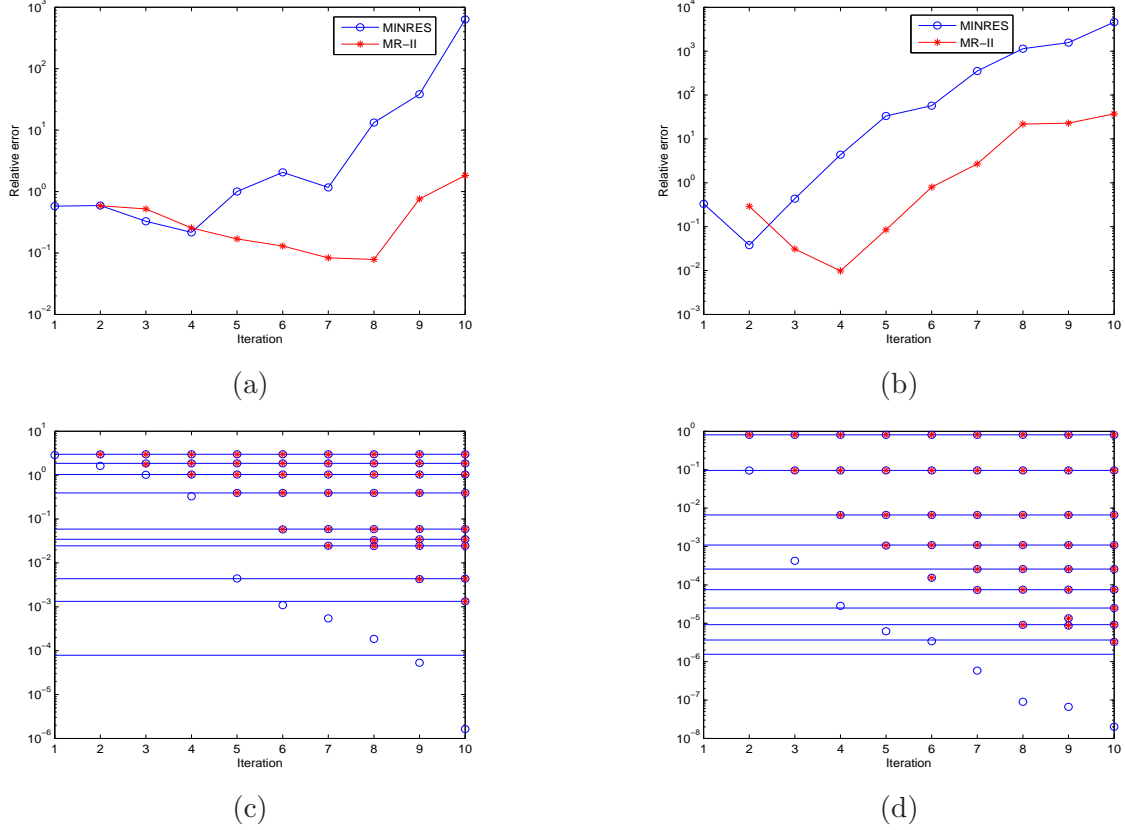


Figure 1: (a)-(b): The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ by MINRES and MR-II; (c)-(d): Plots of the singular values (circles for MINRES, stars for MR-II) of the projected matrices and the ones (solid lines) of A for shaw (left) and foxgood (right).

cannot obtain best possible regularized solutions and has only the partial regularization, independent of the degree of ill-posedness; (ii) the MINRES iterates $\bar{x}^{(k)}$ are always more accurate than the MR-II iterates $x^{(k-1)}$ until the semi-convergence of MINRES.

In this subsection, we only report the results for the noise level $\varepsilon = 10^{-3}$. Results for the other two ε are analogous and thus omitted.

Figures 1 and 2 display numerous curves for severely and moderately ill-posed problems. Clearly, all the MR-II iterates are always more accurate than the MINRES iterates at their respective semi-convergence. This indicates that MINRES has only the partial regularization. As elaborated previously, this is because that a small singular value of the projected matrix \bar{T}_k appears before a regularized solution becomes best, causing that its error does not reach the same error level as that obtained by MR-II. For instance, we see from Figure 1 (a) and (c) that all the singular values of \bar{T}_k in MR-II are excellent approximations to the k large singular values of A in natural order for $k \leq 9$. We see that the semi-convergence of MR-II occurs at iteration $k = 7$. By the comments in the end of Section 2 and the explanations after (5.12), this clearly justifies the full regularization of MR-II, and the best possible regularized solution by MR-II includes *seven* dominant spectral or SVD components. On the other hand, it is clearly seen from Figure 1 (c) that the smallest singular value of \bar{T}_5 in MINRES is smaller than $\sigma_8 = \sigma_{k_0+1}$, making the relative error starts to increase dramatically at iteration 5 and MINRES have only the partial regularization.

Similar phenomena are observed for foxgood, and MR-II has the full regularization with $k_0 = 3$,

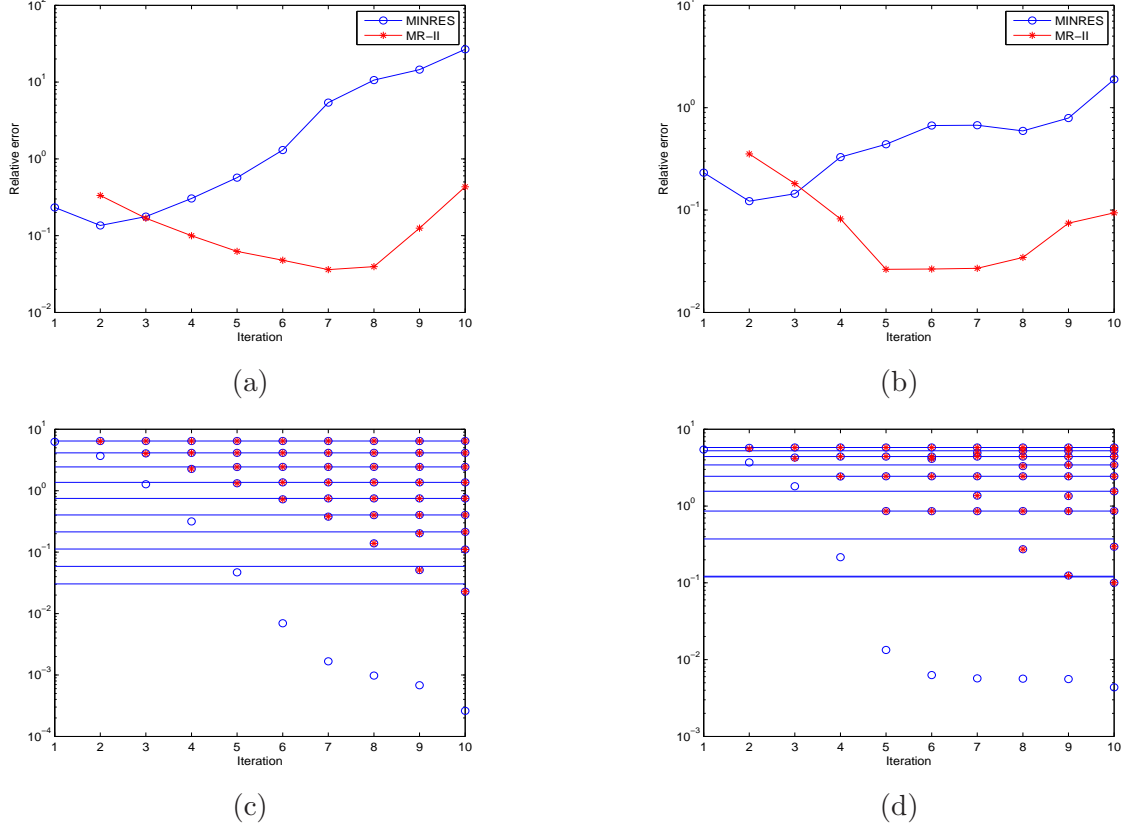


Figure 2: (a)-(b): The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ by MINRES and MR-II; (c)-(d): Plots of the singular values (circles for MINRES, stars for MR-II) of the projected matrices and the ones (solid lines) of A for gravity (left) and phillips (right).

as indicated by Figure 1 (b) and (d), where the smallest singular value of \bar{T}_3 lies between σ_4 and σ_5 and the best iterate $\bar{x}^{(k)}$ by MINRES at semi-convergence is considerably less accurate than the best iterate $x^{(k)}$ by MR-II at semi-convergence, meaning that MINRES has only the partial regularization. We have analogous findings for gravity and phillips, as shown by Figure 2 (b) and (d), which again demonstrate that MR-II has the full regularization but MINRES has only the partial regularization.

As for the mildly ill-posed problem deriv2, we also see from Figure 3 (a) that the relative error obtained by MR-II clearly reaches the lower minimum level than that by MINRES, indicating that MR-II has better regularizing effects than MINRES.

The above experiments have illustrated that MR-II always obtains more accurate regularized solutions than MINRES does for the test severely, moderately and mildly problems. This justifies our theory that MINRES only has the partial regularization, independent of the degree of ill-posedness. Therefore, one must use a hybrid MINRES with some regularization method applied to the projected problems in order to remove the effects of small singular values of \bar{T}_k and improve the accuracy of regularized solutions until a best regularized solution is found.

It is clear from Figures 1-2 and Figure 3 (a) that, for each test problem, the MINRES iterates $\bar{x}^{(k)}$ are more accurate than the corresponding MR-II iterates $x^{(k-1)}$ until the semi-convergence of MINRES. Afterwards, the regularized solutions $\bar{x}^{(k)}$ are deteriorated more and more seriously. This confirms our theory in Section 4, i.e., assertion (ii) in the beginning of this subsection.

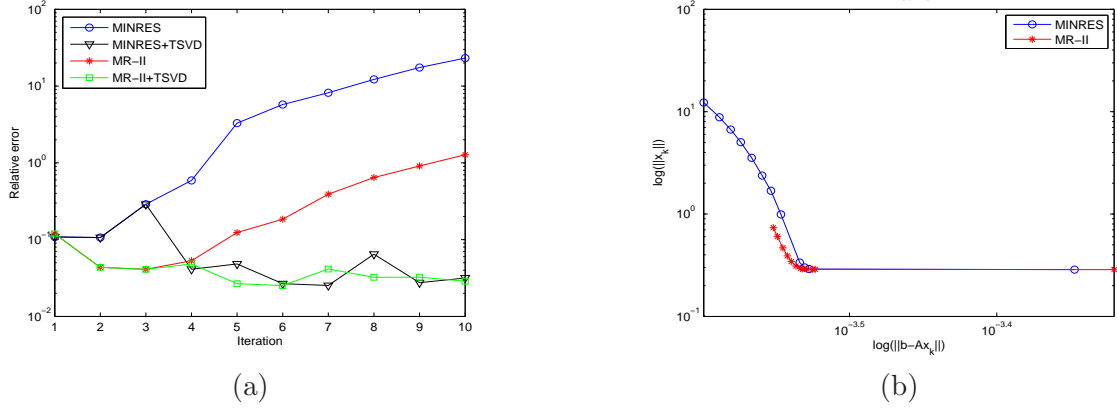


Figure 3: (a): The relative errors $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ by the pure MINRES and MR-II as well as the hybrid MINRES and MR-II; (b): The L-curves of MINRES and MR-II for deriv2.

6.2 The regularizing effects of MR-II, MINRES and their hybrid variants

We first test MR-II, MINRES and their hybrid variants for the mildly ill-posed problem deriv2, and justify our theory that MR-II has only the partial regularization and one must use its hybrid variant to compute a best possible regularized solution.

For deriv2, Figure 3 (a) shows that the relative errors of regularized solutions obtained by the hybrid MINRES and MR-II with the TSVD regularization method applied to the projected problems reach a considerably smaller minimum level than those by MINRES and MR-II themselves. For this problem, before MINRES or MR-II captures all the dominant spectral components needed, a small singular value of \bar{T}_k or T_k appears and starts to deteriorate the regularized solutions. In contrast, their hybrid variants expand Krylov subspaces until enough dominant spectral components are captured and the TSVD regularization method effectively dampens the SVD components corresponding to small singular values of the projected matrices \bar{T}_k by MINRES and T_k by MR-II. For example, we see from Figure 3 (a) that the semi-convergence of MR-II occurs at iteration $k = 3$, which is also observed by the corner of the L-curve depicted by Figure 3 (b). However, as shown by Figure 3 (a), such regularization of MR-II is not enough, and the hybrid MR-II uses a larger six dimensional Krylov subspace $\mathcal{K}_6(A, Ab)$ to improve the solutions and get a best possible regularized solution, whose residual norm is smaller than that obtained by the pure MR-II. After $k = 6$, the regularized solutions almost stabilize with the minimum error as k increases. We observe similar phenomena for MINRES and its hybrid variant, where we find that the relative error by the hybrid MINRES reaches the same minimum level as that by the hybrid MR-II.

Next we test MR-II, MINRES and their hybrid variants for severely and moderately ill-posed problems. We attempt to get more insight into the regularizing effects of MR-II. As a matter of fact, we have already justified the full regularization of MR-II for the four test problems in Section 6.1. In what follows, we will give more details and justifications on the full regularization of MR-II. We show that (i) the relative error obtained by the hybrid MINRES reaches the same minimum level as that by the hybrid MR-II; (ii) MR-II has the full regularization effects: at semi-convergence, the regularized solution by the pure MR-II is as accurate as that by the hybrid MR-II with the TSVD regularization used within projected problems; (iii) MR-II generates near best rank k approximations $Q_{k+1}T_kQ_k^T$ to A , i.e., the relation $\gamma_k \approx \sigma_{k+1} = |\lambda_{k+1}|$ holds with different noise levels. Keep in mind (1.5) and (2.5). This means that $Q_{k+1}T_kQ_k^T$ generated by MR-II plays the same role as A_k , the best rank k approximation to A , so that MR-II has the full regularization.

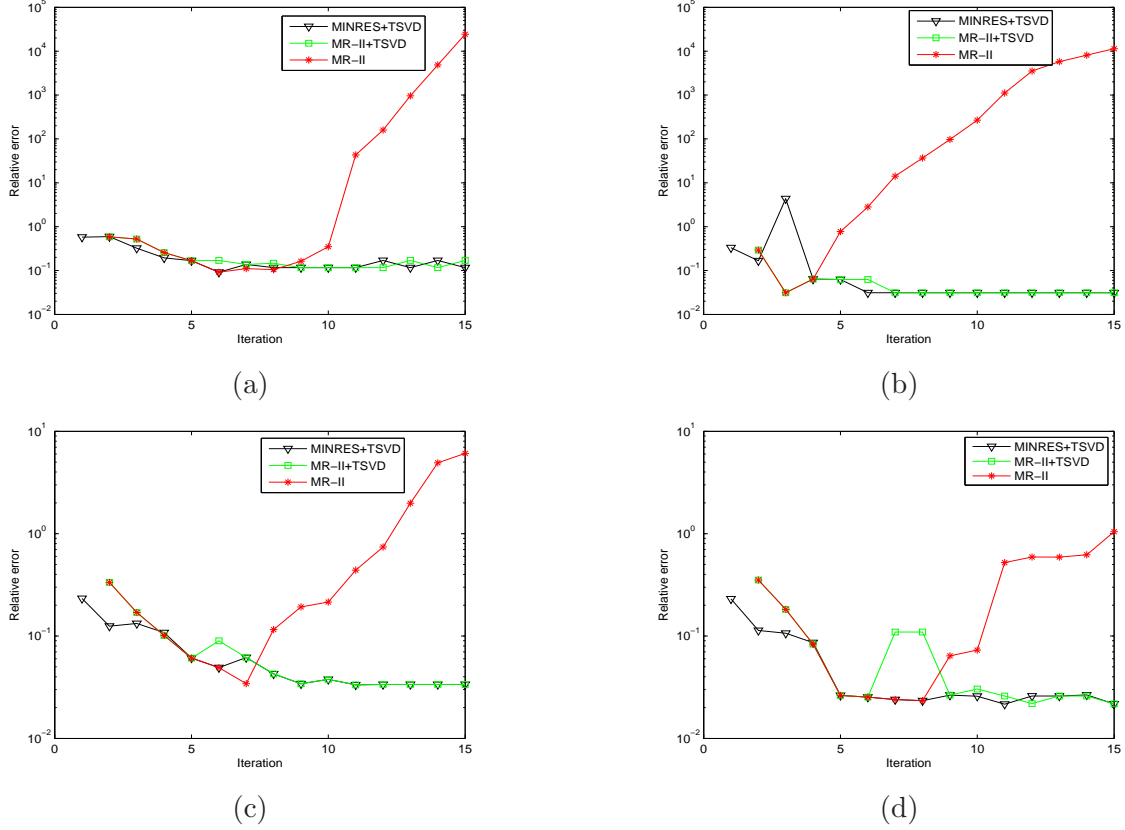


Figure 4: The relative errors $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ by MR-II, and hybrid MR-II and MINRES with additional TSVD regularization for shaw, foxgood, gravity, phillips (from top left to bottom right).

For MR-II and the hybrid MR-II, we observe from Figure 4 that MR-II reaches the same error level as the hybrid MR-II, and the TSVD regularization applied to projected problems simply makes the regularized solutions with the minimum error almost stabilize and does not improve the regularized solution by MR-II at semi-convergence. This justifies the full regularization of MR-II.

Compared with Figures 1–2, we find from Figure 4 that the hybrid MINRES improves on MINRES substantially and the relative errors of iterates by the hybrid MINRES reach the same minimum level as MR-II and the hybrid MR-II. These phenomena again justify our assertion in Section 4 that the hybrid MINRES is necessary, independent of the degree of ill-posedness, and the hybrid MINRES is as effective as the hybrid MR-II.

Figure 5 and Figure 7 display the curves of sequences γ_k with the noise levels $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$, respectively, for the four severely and moderately problems. We see that $\gamma_k \approx \sigma_{k+1} = |\lambda_{k+1}|$, almost independent of noise level ε . We point out that, due to the round-offs in finite precision arithmetic, they level off at the level of ϵ_{mach} when $k = 20$ for shaw, $k = 37$ for foxgood and $k = 50$ for gravity. The results indicate that the $Q_{k+1}T_kQ_k^T$ are near best rank k approximations to A with the approximate accuracy σ_{k+1} so that T_k does not become ill-conditioned before $k \leq k_0$. As a result, the regularized solutions $x^{(k)}$ become increasingly better approximations to x_{true} until iteration k_0 , and they are deteriorated after that iteration. At iteration k_0 , $x^{(k_0)}$ captures the k_0 dominant spectral or equivalent SVD components of A and is a best possible regularized solution, i.e., MR-II has the full regularization for the severely ill-posed problems tested.

Figure 6 and Figure 8 plot the relative errors $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ with different noise levels

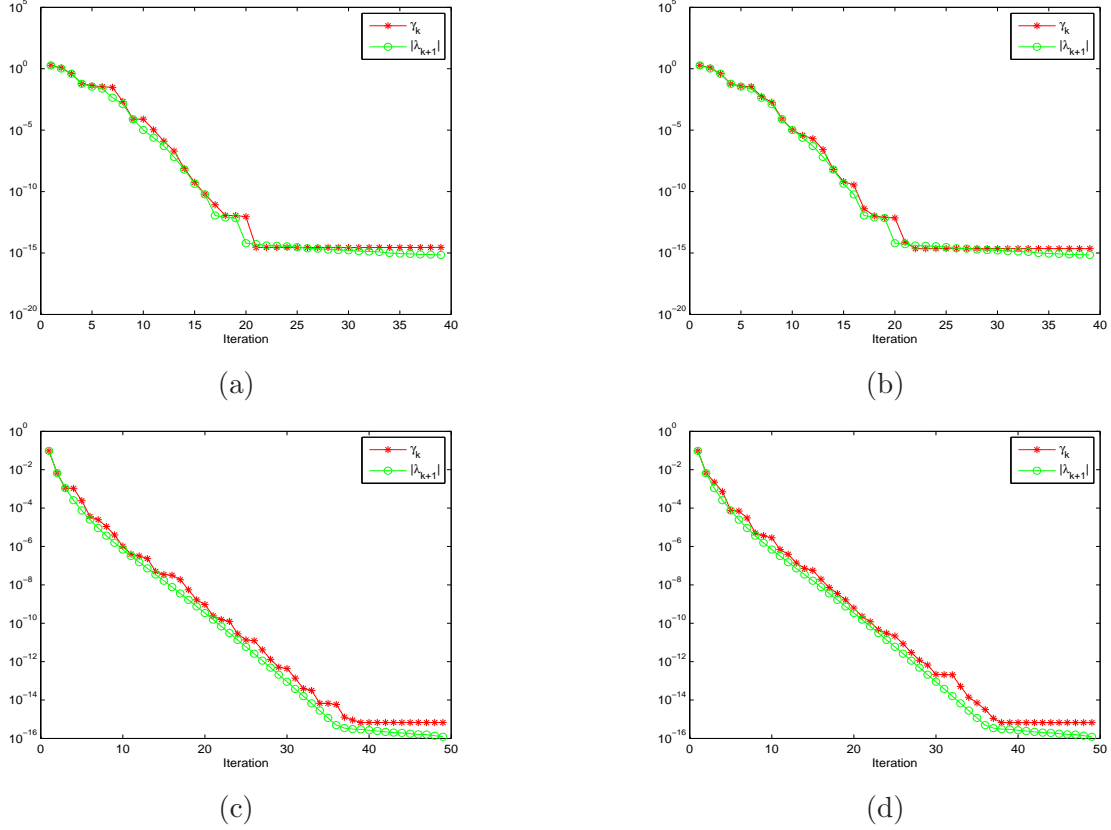
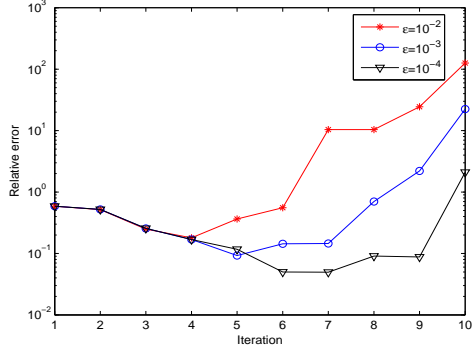


Figure 5: (a)-(b): Plots of decaying behavior of the sequences γ_k and $|\lambda_{k+1}|$ for shaw with $\varepsilon = 10^{-2}$ (left) and $\varepsilon = 10^{-3}$ (right) by MR-II; (c)-(d): Plots of decaying behavior of the sequences γ_k and $|\lambda_{k+1}|$ for foxgood with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right) by MR-II.

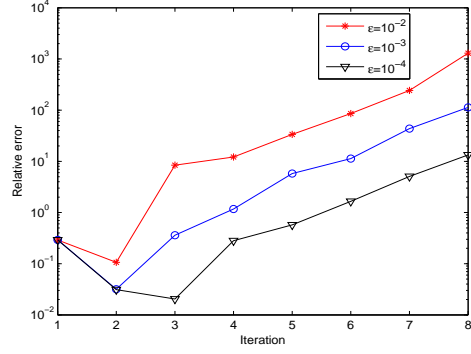
for these four severely and moderately ill-posed problems. For smaller noise levels, MR-II gets more accurate best regularized solutions at cost of more iterations. This is expected since, from (1.4) and $|\lambda_{k_0+1}^{1+\beta}| = |v_{k_0+1}^T \hat{b}| \approx |v_{k_0+1}^T \hat{b}|$, a bigger k_0 is needed for a smaller noise level $\|e\|$. Moreover, MR-II needs more iterations to achieve semi-convergence for moderately ill-posed problems with the same noise level, since σ_j does not decay as fast as that for a severely ill-posed problem.

Figures 9 display the decreasing curves of quantities $|\alpha_{k+1}|$, β_k and σ_k , $k = 2, \dots, n-1$. From Figure 9 (a), we see that, for the severely ill-posed problem shaw, $|\alpha_{k+1}|$ and β_k decrease as fast as σ_k and the three quantities level off at the level of ϵ_{mach} for k no more than 20, and after that these quantities are purely round-offs and not reliable any more. Similar phenomena are also observed for the other two severely ill-posed problems foxgood and gravity, as indicated by Figure 9 (b) and (c). Figure 9 (d) illustrates that β_k decreases as fast as σ_k but $|\alpha_{k+1}|$ decays as fast as σ_k in the first iterations and then considerably faster than σ_k as k increases in the later stage for moderately ill-posed problem phillips.

Finally, we report some comparison results on LSQR, the hybrid LSQR and MR-II, the hybrid MR-II. As already proved in [13], a hybrid LSQR should be used to compute best possible regularized solutions for mildly ill-posed problems. It has also been experimentally justified in [13] that LSQR has the full regularization for severely and moderately ill-posed problems. We have tested LSQR and the hybrid LSQR, and compared their effectiveness and efficiency with MR-II and the hybrid MR-II. We have found that, for each of the severely and moderately ill-posed problems in

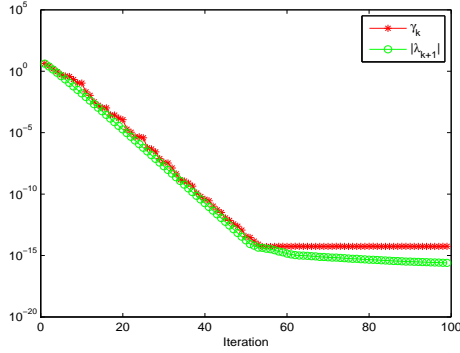


(a)

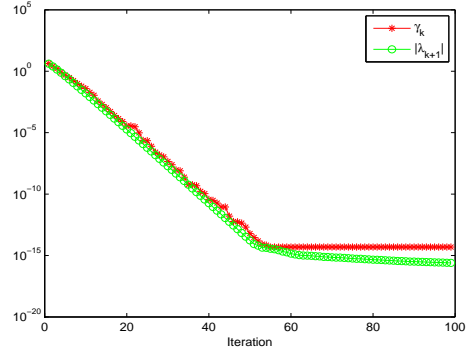


(b)

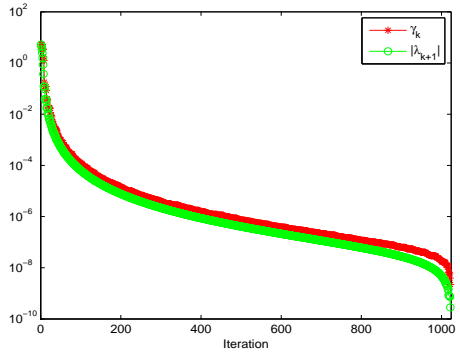
Figure 6: The relative error $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for shaw (left) and foxgood (right) by MR-II.



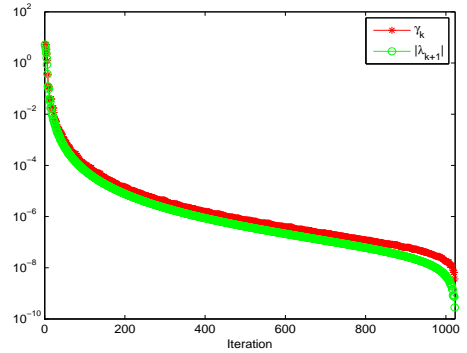
(a)



(b)



(c)



(d)

Figure 7: (a)-(b): Plots of decaying behavior of the sequences γ_k and $|\lambda_{k+1}|$ for gravity with $\varepsilon = 10^{-2}$ (left) and $\varepsilon = 10^{-3}$ (right) by MR-II; (c)-(d): Plots of decaying behavior of the sequences γ_k and $|\lambda_{k+1}|$ for phillips with $\varepsilon = 10^{-3}$ (left) and $\varepsilon = 10^{-4}$ (right) by MR-II.

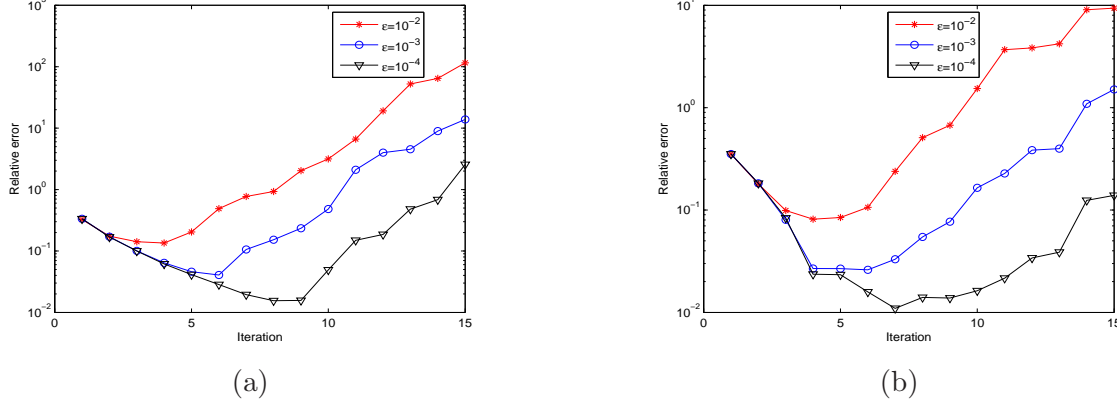


Figure 8: The relative errors $\|x^{(k)} - x_{true}\| / \|x_{true}\|$ with respect to $\varepsilon = 10^{-2}, 10^{-3}, 10^{-4}$ for gravity (left) and phillips (right) by MR-II.

Table 1 and with the same noise level, both the pure MR-II and LSQR obtain the best regularized solutions with almost the same accuracy using almost the same iterations. For deriv2, the hybrid MR-II and LSQR compute the best possible regularized solutions using almost the same iterations. These results tell us two things: (i) As an iterative regularization method, MR-II is as effective as LSQR for an ill-posed problem; (ii) MR-II is twice as efficient as LSQR.

6.3 A 2D image restoration problem

The problem blur is a 2D image deblurring problem and more complex than the other five 1D problems in Table 1. It arises in connection with the degradation of digital images by atmospheric turbulence blur. We use the code `blur(n,band,sigma)` in [9] to generate an $n^2 \times n^2$ matrix A , the true solution x_{true} and noise-free right-hand \hat{b} . The vector x_{true} is a columnwise stacked version of a simple test image, while $\hat{b} = Ax_{true}$ holds for a columnwise stacked version of the blurred image. The blurring matrix A is block Toeplitz with Toeplitz blocks, which has two parameters **band** and **sigma**; the former specifies the half-bandwidth of the Toeplitz blocks, and the latter controls the shape of the Gaussian point spread function and thus the amount of smoothing. We generate a blurred and noisy image $b = \hat{b} + e$ by adding a white noise vector e . The goal is to restore the true image x_{true} from b .

We take $n = 256$ and the relative noise level $\varepsilon = 5 \times 10^{-3}$, giving rise to A with order $n^2 = 65,536$. It is known that the larger the **sigma**, the less ill-posed the problem. Purely for an experimental purpose, we computed all the singular values of a few A with $n^2 \leq 10,000$ using the matlab function `svd`. Since the degree of ill-posedness is the same for different large n^2 , we have deduced from the computed singular values for these matrices A that **band** = 3, **sigma** = 0.7 (the default setting) generates mildly ill-posed problems, while **band** = 7, **sigma** = 2 gives rise to moderately ill-posed problems. We next test MINRES, MR-II and their hybrid variants for these two problems, and verify the regularizing effects similar to the previous mildly and moderately ill-posed problems: deriv2 and phillips.

Figure 10 (a) shows that MINRES and MR-II have the partial regularization for the mildly ill-posed problem blur. The semi-convergence of the two methods occurs at the very first iteration, then regularized solutions are progressively deteriorated, while the hybrid MINRES finds the best possible regularized solution at iteration $k = 9$ and the hybrid MR-II does so at $k = 8$. Moreover, we see that the hybrid MINRES and MR-II reaches the same minimum error level. Figure 10 exhibits

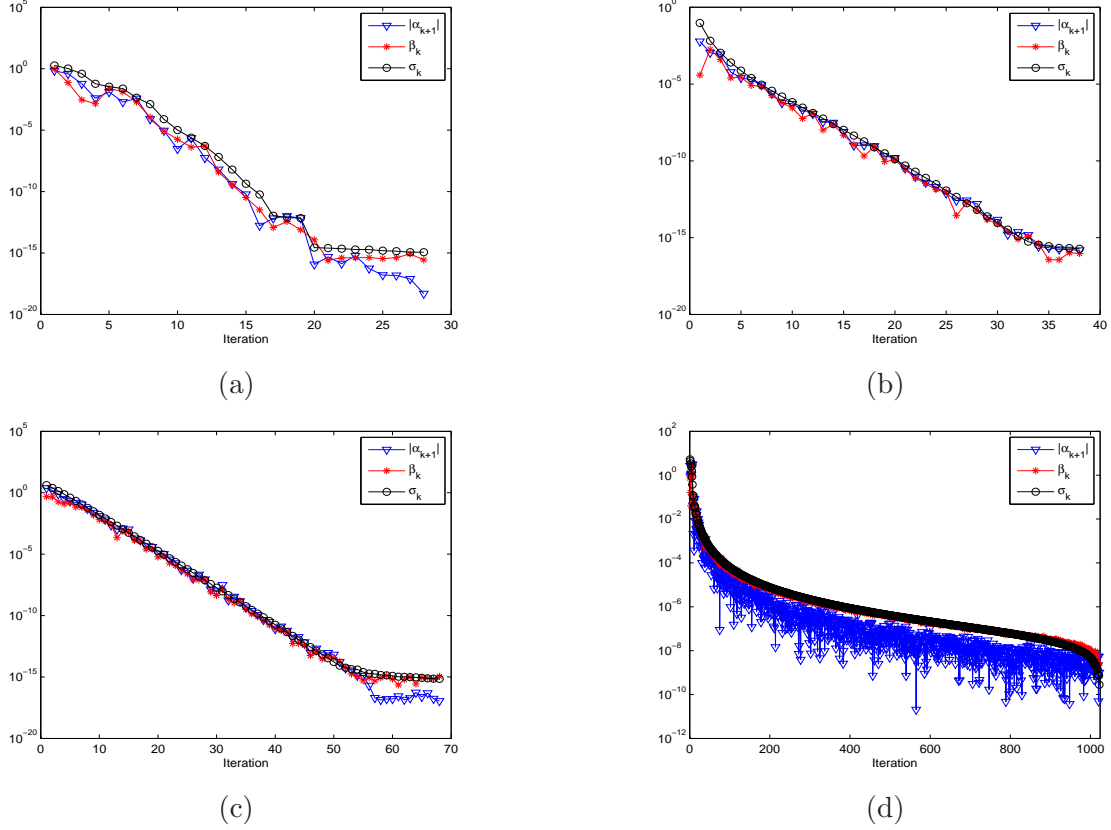


Figure 9: (a)-(d): Plots of decaying behavior of the sequences $|\alpha_{k+1}|$, β_k and σ_k for shaw, foxgood, gravity, phillips (from top left to bottom right) with $\varepsilon = 10^{-3}$ by MR-II.

the restoration performance, where the restored image is chosen by the regularized solution at the iteration where the hybrid MR-II first reaches the minimum error level. We observe from Figure 10 (d) that the outline of original image is restored quite well by the restored image.

From Figure 11 (a), we see that the semi-convergence of MR-II occurs at the first iteration and the regularized solution at this iteration is as accurate as those obtained by the hybrid MR-II and MINRES for the moderately ill-posed problem blur. Therefore, MR-II has the full regularization for this problem. In contrast, MINRES has only the partial regularization because its regularized solution at semi-convergence is much less accurate than that obtained by MR-II. In addition, we observe that the hybrid MINRES and the hybrid MR-II simply make the regularized solutions almost stabilize with the minimum error. Figure 11 (d) exhibits the restored image, which is a good approximation to the original image.

7 Conclusions

For large scale symmetric discrete linear ill-posed problems, MINRES and MR-II are natural alternatives to LSQR. Our theory and experiments have shown that MINRES has only the partial regularization and its hybrid variant is needed to find best possible regularized solutions, independent of the degree of ill-posedness. We have proved that MR-II has better regularizing effects for severely and moderately ill-posed problems than for mildly ill-posed problems, and it generally has only the partial regularization for mildly ill-posed problems. We have shown that although

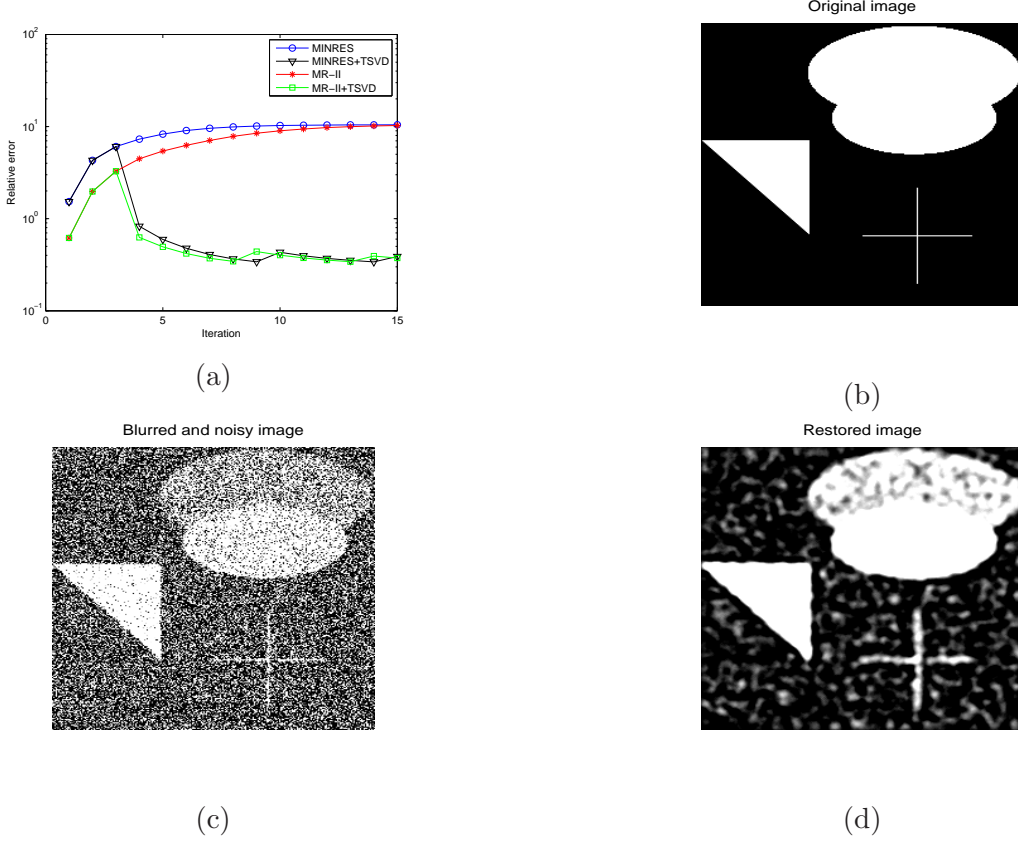


Figure 10: (a): The relative errors $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ by MINRES, hybrid MINRES, MR-II, and hybrid MR-II; (b): The original image; (c): The blurred and noisy image; (d): The restored image with $\text{band} = 3, \text{sigma} = 0.7$.

MR-II is a better regularization method than MINRES, the k th MINRES iterate is always more accurate than the $(k - 1)$ th MR-II iterate until the semi-convergence of MINRES. We have also established estimates for the entries generated by the Lanczos process working on $\mathcal{K}(A, Ab)$, showing how fast they decay. All these results have been confirmed numerically. Remarkably, stronger than our theory predicts, we have numerically demonstrated that MR-II has the full regularization for severely and moderately ill-posed problems and can compute best possible regularized solutions. As a comparison of MR-II and LSQR for a general symmetric ill-posed problem, our theory experiments have indicated that two methods have very similar regularizing effects but MR-II is twice as efficient as LSQR, so do their hybrid variants. Therefore, for a large scale symmetric problem (1.1), MR-II may be preferable to LSQR.

Some problems need to be further considered. As we have seen, more appealing is a sharp estimate for $\|\Delta_k\|$ other than $\|\Delta_k\|_F$. The quantity $\|\sin \Theta(\mathcal{V}_k, \mathcal{V}_k^s)\|$ needs a more subtle analysis and plays a central role in accurately estimating the accuracy γ_k of the rank k approximation generated by the Lanczos process working on $\mathcal{K}(A, Ab)$. As we have elaborated, studying how near γ_k is to σ_{k+1} is a central problem that completely understands the regularizing effects of MR-II. Our bounds in Theorems 5.1–5.2 are less sharp and need to be improved on.

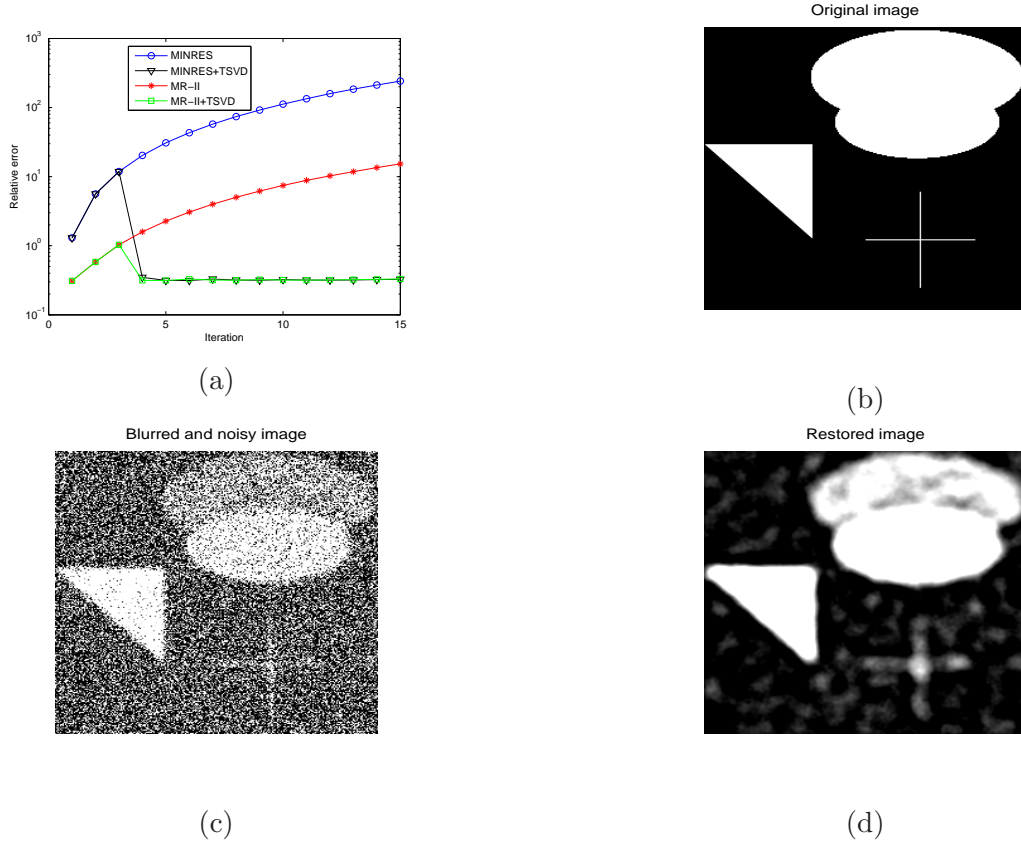


Figure 11: (a): The relative errors $\|x^{(k)} - x_{true}\|/\|x_{true}\|$ with respect to MINRES, hybrid MINRES, MR-II, and hybrid MR-II; (b): The original image; (c): The blurred and noisy image; (d): The restored image with $\text{band} = 7$, $\text{sigma} = 2$.

References

- [1] F. BAUER AND M. A. LUKAS, *Comparing parameter choice for regularization of ill-posed problems*. Math. Comput. Simul., 81 (2011), pp. 1795–1841.
- [2] Å. BJÖRCK, *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, PA, 1996.
- [3] B. FISCHER, M. HANKE AND M. HOCHBRUCK, *A note on conjugate-gradient type methods for indefinite and/or inconsistent linear systems*. Numer. Algor., 11 (1996), pp. 181–187.
- [4] S. GAZZOLA, *Regularization techniques based on Krylov methods for ill-posed linear systems*. Ph. D. thesis, Dept. of Mathematics, University of Padua, Italy, 2014.
- [5] M. HANKE, *Conjugate Gradient Type Methods for Ill-Posed Problems*. Longman, Essex, 1995.
- [6] M. HANKE, *On Lanczos based methods for the regularization of discrete ill-posed problems*. BIT Numer. Math., 41 (2001), pp. 1008–1018.
- [7] M. HANKE AND J. G. NAGY, *Restoration of atmospherically blurred images by symmetric indefinite conjugate gradient techniques*. Inverse Probl., 12 (1996), pp. 157–173.
- [8] P. C. HANSEN, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, Philadelphia, PA, 1998.
- [9] P. C. HANSEN, *Regularization tools version 4.0 for Matlab 7.3*. Numer. Algor., 46 (2007), pp. 189–194.
- [10] P. C. HANSEN, *Discrete Inverse Problems: Insight and Algorithms*. SIAM, Philadelphia, PA, 2010.

- [11] P. C. HANSEN AND T. K. JENSEN, *Noise propagation in regularizing iterations for image deblurring*. Electron. Trans. Numer. Anal., 31 (2008), pp. 204–220.
- [12] B. HOFMANN, *Regularization for Applied Inverse and Ill-posed Problems*. Teubner, Stuttgart, Germany, 1986.
- [13] Y. HUANG AND Z. JIA, *Some results on regularization of LSQR for large-scale discrete ill-posed problems*. arXiv: math.NA/1503.01864v1, 2015.
- [14] T. K. JENSEN AND P. C. HANSEN, *Iterative regularization with minimum-residual methods*. BIT Numer. Math., 47 (2007), pp. 103–120.
- [15] M. E. KILMER AND D. P. O’LEARY, *Choosing regularization parameters in iterative methods for ill-posed problems*. SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221.
- [16] M. E. KILMER AND G. W. STEWART, *Iterative regularization and minres*. SIAM J. Matrix Anal. Appl., 21 (1999), pp. 613–628.
- [17] A. NEUMAN, L. REICHEL AND H. SADOK, *Algorithms for range restricted iterative methods for linear discrete ill-posed problems*. Numer. Algor., 59 (2012), pp. 325–331.
- [18] C. C. PAIGE, B. N. PARLETT AND H. A. VAN DER VORST, *Approximate solutions and eigenvalue bounds from Krylov subspaces*. Numer. Linear Algebra Appl., 2 (1995), pp. 115–133.
- [19] C. C. PAIGE AND M. A. SAUNDERS, *LSQR: an algorithm for sparse linear equations and sparse least squares*. ACM Trans. Math. Softs., 8 (1982), pp. 43–71.
- [20] C. C. PAIGE AND M. A. SAUNDERS, *Solution of sparse indefinite systems of linear equations*. SIAM. J. Numer. Anal., 12 (1975), pp. 617–629.
- [21] L. REICHEL AND G. RODRIGUEZ, *Old and new parameter choice rules for discrete ill-posed problems*. Numer. Algor., 63 (2013), pp. 65–87.
- [22] G. W. STEWART, *Matrix Algorithms. Volume II: Eigensystems*, SIAM, Philadelphia, PA, 2001.